Sensitivity Analysis of Transient Storage Model Predictions – Narew National Park Case Study

Marzena OSUCH, Adam KICZKO, Jarosław NAPIÓRKOWSKI
and Renata ROMANOWICZ

Institute of Geophysics Polish Academy of Sciences
Ks. Janusza 64, 01-452 Warsaw, Poland
e-mail: marz@igf.edu.pl

Abstract

A fluorescent dye-tracer study was performed under steady-state flow conditions on a 16.8 km reach of an anastomosing section of the Upper Narew River in order to better understand the solute-transport processes in a wetland system. The procedure consisted of the instantaneous injection of a known quantity of the solution of Rhodamine WT into a stream and observation of the variation in concentration of the tracer as it moved downstream. The paper describes the sensitivity analysis of a transient storage model applied to the experimental data. Special emphasis is given to ecologically related measures, such as estimates of the peak of tracer concentrations at cross-sections along the river and the length of time when the concentrations exceed specified threshold.

1. Introduction

The present study has been motivated by the need for estimating the risk involved with the spread of pollutants in a unique river system situated within the Narew National Park.

The advection-dispersion model with dead zones that can adequately describe the process of transport of pollutants in a single-channel river with multiple storages (Bencala and Walters 1983, Rowiński *et al.* 2003a, b) was applied to the data from a dye-tracer experiment performed on a 16.8 km reach of an anastomosing section of the Upper Narew River.

The chosen model is deterministic, i.e., it assumes that observations are without errors and the model structure perfectly describes the process of transport. In order to take into account the model and observation errors, an uncertainty analysis is required. Following the discussion presented by Romanowicz and Macdonald (2005), the first step of the uncertainty analysis consists of a sensitivity analysis of the model output followed by an estimation of parameter uncertainty conditioned on the available ob-

servations. The uncertainty of model predictions is estimated on the basis of parametric conditional uncertainty. In this paper we discuss the first step of an uncertainty analysis which is the application of the Global Sensitivity Analysis (GSA), introduced by Archer *et al*. (1997). This concerns the relationship between the parameters and supports the choice of parameters which contribute the most to the model predictive uncertainty. The influence of different model parameters on ecologically-related measures such as maximum concentrations at cross-sections along the river and time periods with concentration exceeding the ecologically safe threshold is also investigated.

2. Distributed transient storage model

The present paper is based on a tracer test performed on a unique multi-channel system on the Narew River reach within the Narew National Park in northeast Poland (Fig. 1). A detailed description of the experiment is presented in Rowiński (2003a, b). The One-dimensional Transport with Inflow and Storage model (OTIS) introduced by Bencala and Walters (1983) was applied here. The OTIS model is formed by writing mass balance equations for two conceptual areas, the stream channel and the storage zone. The stream channel is defined as that portion of the stream in which advection and dispersion are the dominant transport mechanisms. The storage zone is defined as the portion of the stream that contributes to transient storage, i.e. stagnant pockets of
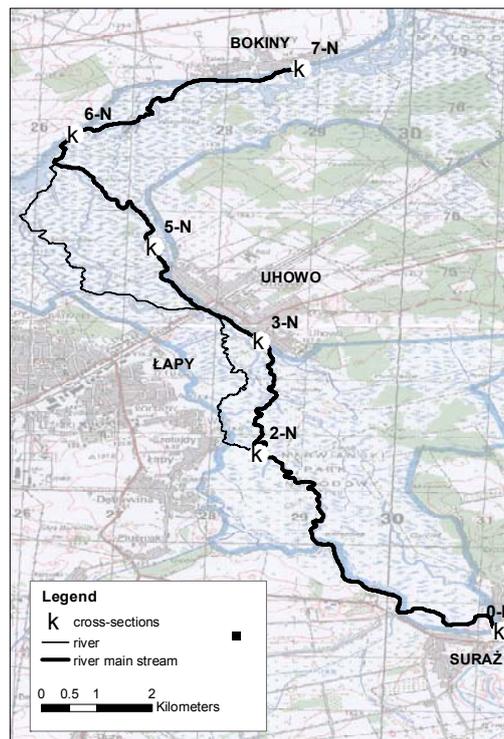
Fig. 1. Map of the experimental reach of Upper Narew River.

water and porous areas of the streambed. Water in the storage zone is considered immobile relative to water in the stream channel. The exchange of solute mass between the stream channel and the storage zone is modelled as a first-order mass transfer process.

Since it is not possible to estimate solute transport parameters reliably from hydraulic variables and channel characteristics, application of the transient storage model requires the estimation of model parameters for each particular river reach (2N-3N, 3N-5N, 5N-6N and 6N-7N; Fig. 1) based on data from tracer experiment and measurements of lateral inflow and discharge. Estimation of model parameters, namely the coefficient of longitudinal dispersion D, the main channel cross-sectional area A, the storage zone cross-sectional area A_S, and the exchange coefficient α was performed by minimizing the residuals between the simulated and observed concentrations. A general least square objective function and Nealder-Mead minimization algorithm was used.

The results of the estimation procedure are given in Table 1. They are similar to those obtained by Rowiński *et al*. (2004) for a similar model but different numerical scheme.

Table 1

Parameters of transient storage models

Parameters	Sections			
	2N-3N	3N-5N	5N-6N	6N-7N
D [m2/s]	10.31	1.65	6.96	1.59
A [m2]	9.71	34.70	11.29	25.02
A_S [m2]	6.13	22.62	4.46	7.05
α [1/s]	0.482* 10-5	1.7863* 10-5	1.2913* 10-5	6.5701* 10-5

Note that values of the parameters differ from reach to reach. These big differences in parameter values result from the high variability of geometric and hydraulic conditions between the reaches.

The ordinary least squares criterion is used to ensure that the model reproduces adequately the observed transport processes. However, from the point of view of the ecology of the wetlands, for each cross section i, the estimation of maximum concentration of the tracer $C_{max,j} = \max(C_i(t))$ and time periods $T_{thr,i}(C_{thr})$, during which a safe threshold level of concentration C_{thr} is exceeded, are very important.

3. Sensitivity analysis

Generally, the sensitivity analysis consists of an evaluation of the relation between input and output variations. In this assessment we have used the variance based Global Sensitivity Analysis approach introduced by Archer *et al*. (1997). According to this method, the whole set of model parameters acquired from the Monte Carlo sampling is

analysed simultaneously and there is no restriction on the monotonicity or additivity of the model. Therefore this approach is suitable for over- parameterized, nonlinear, spatially distributed models.

According to this methodology, the variance of an output Y depending on the variable input set X_i, is based on estimating the fractional contribution of each input factor to the variance of the model output. In order to calculate the sensitivity indices for each factor, the total variance V of the model output is decomposed as:

$$V = \sum_i V_i + \sum_{i<j} V_{ij} + \sum_{i<j<m} V_{ijm} + \ldots + V_{12\ldots k} \tag{1}$$

where

$$V_i = V(E(Y|X_i = x_i^*)) \tag{2}$$

$$V_{ij} = V(E(Y|X_i = x_i^*, X_j = x_j^*)) - V(E(Y|X_i = x_i^*)) - V(E(Y|X_j = x_j^*)) \tag{3}$$

In above formulas, Y denotes the output variable, X_i denotes an input factor, $E(Y|X_i = x_i^*)$ denotes the expectation of Y conditional on X_i having fixed value x_i and others are normally varying. The decomposition is unique if the X_i are independent from each other.

The direct sensitivity of output Y to the input X_i, represents the Sobol first order sensitivity index S_i which takes the following form:

$$S_i = \frac{V[E(Y|X_i = x_i^*)]}{V(Y)} \tag{4}$$

where $V(E(V(Y|X_i = x_i^*))$ is the variance of estimated Y output where X_i parameters are fully fixed and others are normally varying. First order sensitivity index represents the average output variance reduction that can be achieved when X_i becomes fully known and is fixed.

The model sensitivity to the interactions among subsets of factors, the so-called higher order effects, are investigated with the use of the Sobol total sensitivity indices: S_{Ti}. They represent the whole range of interactions which involve X_i and are defined as:

$$S_{Ti} = \frac{E(V(Y|X_{-i} = x_{-i}^*)]}{V(Y)} \tag{5}$$

where $E(V(Y|X_{-i} = x_{-i}^*))$ is estimated variance in case when all parameters are fixed, except X_i which is varying.

The use of total sensitivity indices is advantageous, because there is no need for the evaluation of a single indicator for every possible parameter combination. On the basis of these two indices, S_i and S_{Ti}, it is possible to trace the significance of each model parameter in an efficient way.

4. Discussion

Sobol first order and Sobol total order sensitivity indices for the parameters of the OTIS model predictions are shown in Fig. 2. The order of parameters for any particular river reach is the same. The main channel cross-sectional area A has the largest influence on the output. The exchange coefficient (α) and the storage zone area (A_S) have smaller values of indices indicating smaller influence on the model output and smaller identifiability of these parameters. The lowest values of Sobol first and total sensitivity indices are obtained for dispersion coefficient (D).
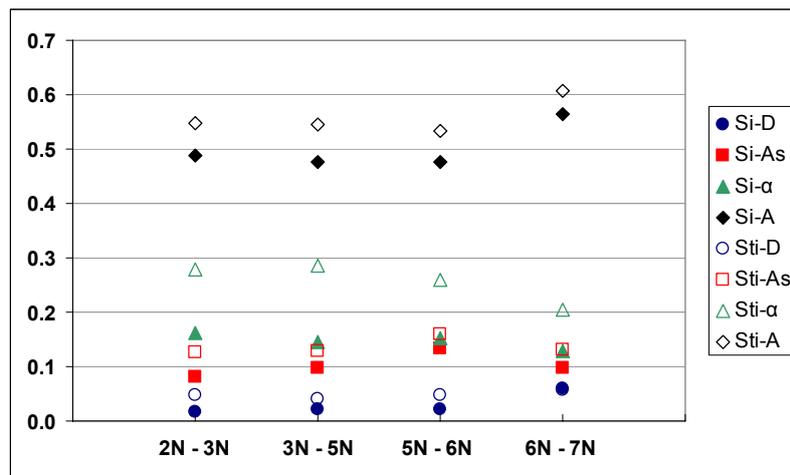
Fig. 2. Sobol first and total order sensitivity indices for OTIS model predictions for all cross-sections. Circles, squares, triangles and diamonds denote dispersion coefficient D, area of storage zone A_S, exchange coefficient α, and area of the main channel A, respectively.

Table 2

Results of the GSA sensitivity analysis on maximum concentration

River reach	S_i				S_{Ti}			
	D	A_S	α	A	D	A_S	α	A
N2-N3	0.028	0.072	0.164	0.637	0.049	0.110	0.208	0.633
N3-N5	0.028	0.068	0.143	0.640	0.040	0.108	0.206	0.645
N5-N6	0.026	0.122	0.186	0.615	0.039	0.148	0.222	0.590
N6-N7	0.047	0.130	0.173	0.578	0.053	0.171	0.211	0.565

The sensitivity indices of the first ecologically related measure, maximum concentration of the tracer, on model parameter are shown in Table 2. The values of sensitivity indices are similar for all four analyzed river reaches and they resemble the results obtained for model predictions. There is a relationship between the area of the main channel and the values of maximum concentration. In the case of the exchange coeffi-

cient and the storage zone area this relationship is weaker and the dispersion coefficient shows the smallest influence on maximum concentrations.

Results for the "over the threshold" periods depend on the threshold value. Figure 3 presents the first and total order sensitivity indices of the OTIS parameter variations as a function of the threshold value. It is interesting to note that the sensitivity of the "over the threshold" period for small and large threshold values shows different behaviour, shown in Fig. 3 as multiple minima/maxima. This behaviour is the result of two different processes. One is the direct influence of parameters on different parts of the dynamic response of the system. The other is the dependence of the maximum peak concentration at each cross-section on the parameter values, i.e., for high threshold values, the number of realisations with a non-zero "over the threshold" period decreases.
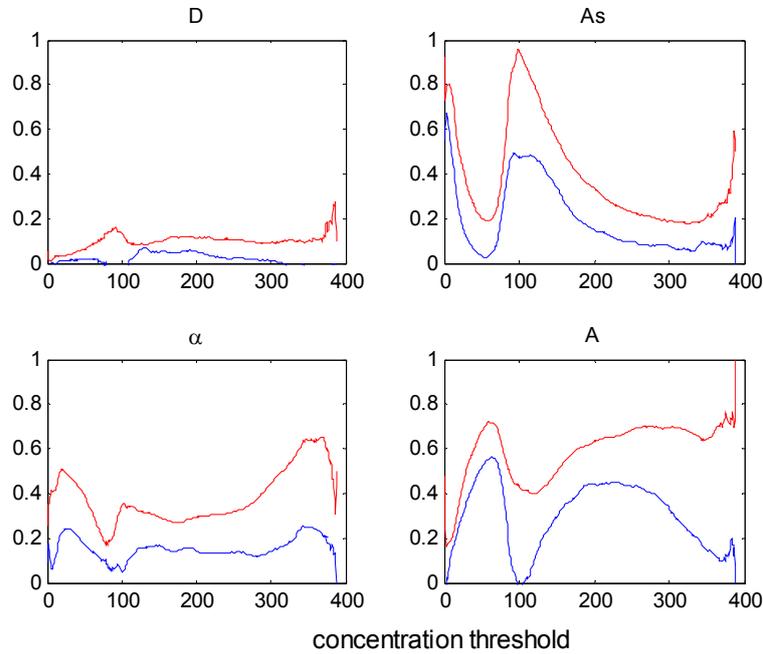
Fig. 3. Sensitivity indices for "over the threshold" period to OTIS parameters variations as a function of the threshold value. Blue and red lines denote Sobol first and total order sensitivity indices, respectively.

In order to explain this behaviour, we shall analyse the projections of the response surface for the parameter A_S for four different values of the threshold, 10, 60, 100 and 200 ppb (Fig. 4a-d). For small threshold values (Fig. 4a) the storage zone area influences the number of time periods over the threshold due to its influence on the tails of the dynamic response of the system (Wagener *et al*. 2002). This influence decreases with an increase of the threshold value, resulting in the minimum index value at the threshold of about 60 ppb (Fig. 4b). Above this threshold, due to the dependence of maximum concentration on the storage zone area A_S for higher values of this parame-

ter, there is an increasing number of realizations for which the threshold concentration of 100 ppb is not reached (Fig. 4c).
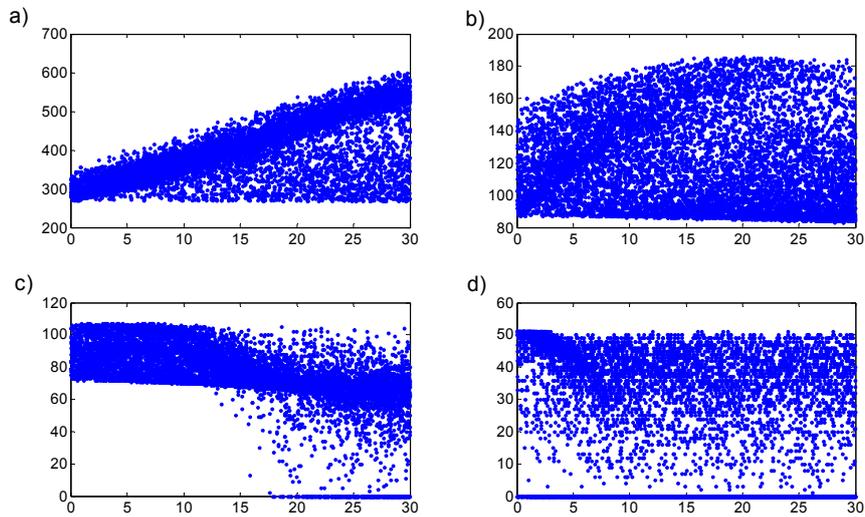
Fig. 4. Sensitivity analyses for the "over the threshold" period for the 6N-7N river reach. Dotted plots a, b, c and d show the projection of the response surface (number of time steps with concentration over the threshold) into the parameter A_S dimension for four threshold values: 10, 60, 100, and 200 ppb, respectively.
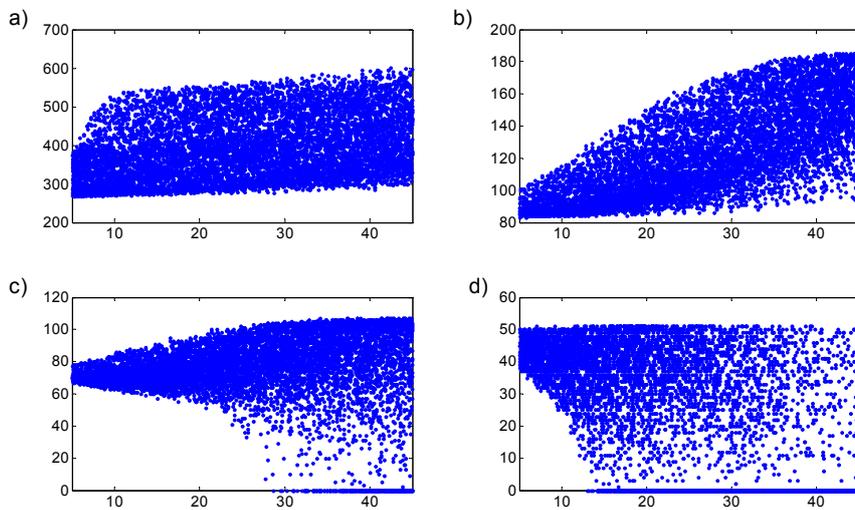
Fig. 5. Sensitivity analyses for the "over the threshold" period for the 6N-7N river reach. Dotted plots a, b, c and d show the projection of the response surface (number of time steps with concentration over the threshold) into the parameter A dimension for four threshold values: 10, 60, 100, and 200 ppb, respectively.

As a result, the number of realizations with decreasing or equal to zero "over the threshold" periods increases, giving a rise of the sensitivity index for this parameter. With further increase of the threshold value, the number of realizations with "over the threshold" period stabilizes, as there are zero-length "over the threshold" periods over the whole parameter range (Fig. 4d). It is interesting to note nearly opposite relationship for parameter A (main channel cross-sectional area), shown in detail in Figs. 5 a-d for threshold values equal to 10, 60, 100, and 200 ppb, respectively. This parameter influences higher parts of the dynamic response of the system giving a rise of the sensitivity index with an increase of the threshold value (Fig. 5a and b). With further increase of the threshold values, zero periods appear that counteract the increase of the number of over the threshold periods, thus decreasing the sensitivity index (Fig. 5c). This influence is limited to the higher values of that parameter, which results in subsequent rise of the sensitivity index for values of the threshold higher than 100 ppb (Fig. 5d).

5. Conclusions

The results of tracer experiments can give an insight into the processes of transport of pollutants in the complex River Narew system. However, the uncertainty of tracer observations and model parameters due to the unavoidable simplifications in process description should be taken into account. In this paper we show that deterministic model predictions span the whole range of values and will differ depending on the model output. We applied sensitivity analysis to define the most sensitive parameters and their ranges. Apart from the time trajectory, maximum concentrations and the length of time period with concentrations exceeding the specified threshold were also used as ecologically related model outputs. In particular, the sensitivity analysis of the latter shows an interesting relationship for threshold values below 200 ppb. The results of this analysis can be used to specify of the best parameter ranges and their prior distributions for the evaluation of predictive model uncertainty using the Generalised Likelihood Uncertainty Analysis (GLUE) of Beven and Binley (1992).

References

Archer, G., A. Saltelli and I.M. Sobol, 1997, Sensitivity measures, anova-like techniques and the use of bootstrap, *Journal of Statistical Computation and Simulation*, 58:99-120.

Bencala, K.E., and R.A. Walters, 1983, Simulation of solute transport in a mountain pool-and-riffle stream: a transient storage model, *Water Resources Research*, 19(3), 718-724.

Beven, K.J., and A. Binley, 1992, The future of distributed models: model calibration and uncertainty prediction, *Hydrological Processes*, 6, 279-298.

Romanowicz, R.J., and R. Macdonald, 2005, Modelling uncertainty and variability in environmental systems, *Acta Geophysica Polonica*, 53, 401-417.

Rowiński, P.M., J. Napiórkowski and J. Szkutnicki, 2003, Transport of passive admixture in a multi-channel river system- the Upper Narew case study. Part 1. Hydrological survey, *Ecology and Hydrobiology*, 3, 371-379.

Rowiński, P.M., J. Napiórkowski and A. Owczarczyk, 2003, Transport of passive admixture in a multi-channel river system- the Upper Narew case study. Part 2. Application of dye tracer method, *Ecology and Hydrobiology*, 3, 381-388.

Rowiński, P.M., J. Napiórkowski and T. Dysarz, 2004, Transport pollution in the rivers flowing thought wetland areas. **In:** *Model application for wetlands hydrology and hydraulics*, Warsaw Agricultural University Press.

Wagener, T., L.A. Camacho and H.S. Wheater, 2002, Dynamic identifiability analysis of the transient storage model for solute transport in rivers, *Journal of Hydroinformatics*, 4(3), 199-211.